ANOVA

Analysis of Variance

Chapter 16

ANOVA

- A procedure for comparing more than two groups
  - **independent variable**: smoking status
    - *non-smoking*
    - *one pack a day*
    - *> two packs a day*
  - **dependent variable**: number of coughs per day
- k = number of conditions (in this case, 3)

One-Way ANOVA

- One-Way ANOVA has one independent variable (1 *factor*) with > 2 *conditions*
  - conditions = levels = treatments
  - e.g., for a brand of cola factor, the levels are:
    - Coke, Pepsi, RC Cola
- Independent variables = factors

Two-Way ANOVA

- Two-Way ANOVA has 2 independent variables (factors)
  - each can have multiple conditions

Example

- Two Independent Variables (IV's)
  - IV1: Brand; and IV2: Calories
  - Three levels of Brand:
    - Coke, Pepsi, RC Cola
  - Two levels of Calories:
    - Regular, Diet

When to use ANOVA

- One-way ANOVA: you have more than two levels (conditions) of a single IV
  - EXAMPLE: studying effectiveness of three types of pain reliever
    - *aspirin* vs. *tylenol* vs. *ibuprofen*
- Two-way ANOVA: you have more than one IV (factor)
  - EXAMPLE: studying pain relief based on pain reliever and type of pain
    - Factor A: Pain reliever (*aspirin* vs. *tylenol*)
    - Factor B: type of pain (*headache* vs. *back pain*)

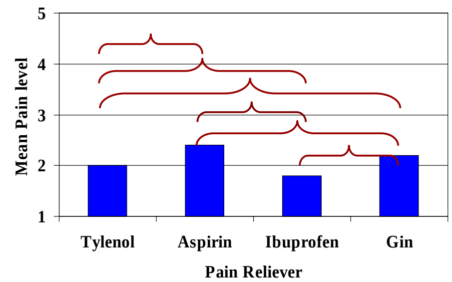
ANOVA

- When a factor uses independent samples in all conditions, it is called a between-subjects factor
  - between-subjects ANOVA
- When a factor uses related samples in all conditions, it is called a within-subjects factor
  - within-subjects ANOVA
  - *PASW: referred to as repeated measures*

ANOVA & PASW

	2 samples	2 or more samples
Independent Samples	Independent Samples *t*-test	Between Subjects ANOVA
Related Samples	Paired Samples *t*-test	Repeated Measures ANOVA

Why bother with ANOVA?



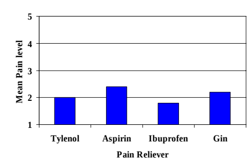
Would require six *t*-tests, each with an associated Type I (false alarm) error rate.

Familywise error rate

- Overall probability of making a Type I (false alarm) error somewhere in an experiment
- One *t*-test,
  - familywise error rate is equal to α (e.g., .05)
- Multiple *t*-tests
  - result in a familywise error rate much larger than the α we selected
- ANOVA keeps the familywise error rate equal to α

Post-hoc Tests

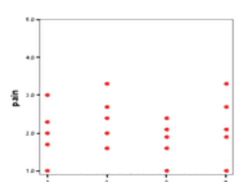
- If the ANOVA is significant
  - *at least* one significant difference between conditions
- In that case, we follow the ANOVA with post-hoc tests that compare two conditions at a time
  - post-hoc comparisons identify the specific significant differences between each pair

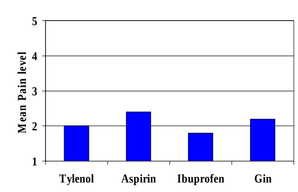


ANOVA Assumptions

- Homogeneity of variance
  - $\sigma^2_1 = \sigma^2_2 = \sigma^2_3 = \sigma^2_4 = \sigma^2_5$
- Normality
  - scores in each population are normally distributed

Partitioning Variance





Partitioning Variance

Total Variance in Scores

Variance Within Groups (error)	**Variance Between Groups**
(chance variance)	***(systematic variance: treatment effect)***

Partitioning Variance

Total Variance in Scores

$\boldsymbol{MS}_{\text{error}}$	$\boldsymbol{MS}_{\text{group}}$

- MS = Mean Square; short for "mean squared deviation"
- Similar to variance (s_x^2)

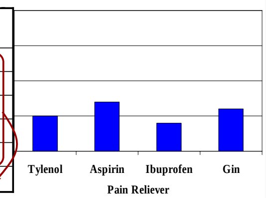
$$t_{\text{obt}} = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$$

$$= \frac{\text{difference between sample means}}{\text{difference expected by chance (standard error)}} = \frac{\text{systematic variance}}{\text{chance variance}}$$

$$F_{obt} = \frac{\text{variance between sample means}}{\text{variance expected by chance (error)}} = \frac{\text{systematic variance}}{\text{chance variance}}$$

- MS_{error} is an estimate of the variability as measured by differences within the conditions
  - sometimes called the within group variance or the error term
  - chance variance (random error + individual differences)

	Tylenol	Aspirin	Ibuprofen	Gin
	3	1.6	2.1	1
	2.3	2.7	1.6	3.3
	1.7	2.4	2.4	1.9
	1	2	1.9	2.7
	2	3.3	1	2.1
Mean:	**2**	**2.4**	**1.8**	**2.2**



MS_{error} = error variance (within groups)

- MS_{group} is an estimate of the differences in scores that occurs between the levels in a factor
  - also called MS_{between}
  - Treatment effect (systematic variance)

	Tylenol	Aspirin	Ibuprofen	Gin
	3	1.6	2.1	1
	2.3	2.7	1.6	3.3
	1.7	2.4	2.4	1.9
	1	2	1.9	2.7
	2	3.3	1	2.1
$\bar{X}$:	**2**	**2.4**	**1.8**	**2.2**

Overall $\bar{X} = 2.1$

MS_{group} = variance between groups

Total Variance
(variability among all the scores in the data set)

Between Groups Variance
1. Treatment effect (systematic)
2. Chance *(random error + individual differences)*

Error Variance (within groups)
1. Chance *(random error + individual differences)*

$$F\text{-}Ratio = \frac{\text{between group variance}}{\text{error variance (within groups)}}$$

$$F\text{-}Ratio = \frac{\textit{Treatment effect + Chance}}{\textit{Chance}}$$

- When H_0 is TRUE (there is no treatment effect):

$$F = \frac{0 + \textit{Chance}}{\textit{Chance}} \cong 1$$

- When H_0 is FALSE (there is a treatment effect):

$$F = \frac{\textit{Treatment effect + Chance}}{\textit{Chance}} > 1$$

- In ANOVA, variance = Mean Square (*MS*)

$$F\text{-}Ratio = \frac{\text{between group variance}}{\text{error variance (within groups)}} = \frac{MS_{\text{group}}}{MS_{\text{error}}}$$

Signal-to Noise Ratio

- ANOVA is about looking at the *signal* relative to *noise*
- MS_{group} is the *signal*
- MS_{error} is the *noise*
- We want to see if the between-group variance (signal), is comparable to the within-group variance (noise)

Logic Behind ANOVA

- If there is no true difference between groups at the population level:
  - the only differences we get between groups in the sample should be due to error.
  - if that's the case, differences between groups should be about the same as differences among individual scores within groups (error).
  - MS_{group} and MS_{error} will be about the same.

Logic Behind ANOVA

- If there are true differences between groups:
  - variance between groups will exceed error variance (within groups)
  - F_{obt} will be much greater than 1

$$F_{\text{obt}} = \frac{MS_{\text{group}}}{MS_{\text{error}}}$$

- F_{obt} can also deviate from 1 by chance alone
  - we need a sampling distribution to tell how high F_{obt} needs to be before we reject the H_0
  - compare F_{obt} to critical value (e.g., $F_{.05}$)

Logic Behind ANOVA

- The critical value ($F_{.05}$) depends on
  - *degrees of freedom*
    - between groups: $df_{\text{group}} = k - 1$
    - error (within): $df_{\text{error}} = k\,(n - 1)$
    - Total: $df_{\text{total}} = N - 1$
  - alpha (α)
    - e.g., .05, .01

ANOVA Example: Cell phones

Research Question:

- Is your reaction time when driving slowed by a cell phone? Does it matter if it's a hands-free phone?

- Twelve participants went into a driving simulator.
  1. A random subset of 4 drove while listening to the radio (control group).
  2. Another 4 drove while talking on a cell phone.
  3. Remaining 4 drove while talking on a hands-free cell phone.
- Every so often, participants would approach a traffic light that was turning red. The time it took for participants to hit the breaks was measured.

A 6 Step Program for Hypothesis Testing

1. State your research question
2. Choose a statistical test
3. Select alpha which determines the critical value ($F_{.05}$)
4. State your statistical hypotheses (as equations)
5. Collect data and calculate test statistic (F_{obt})
6. Interpret results in terms of hypothesis
   Report results
   Explain in plain language

A 6 Step Program for Hypothesis Testing

1. State your research question
   - Is your reaction time when driving influenced by cell-phone usage?
2. Choose a statistical test
   - three levels of a single independent variable (cell; hands-free; control)

     → One-Way ANOVA, between subjects

3. Select α, which determines the critical value

- α = .05 in this case

- See F-tables (page 543 in the Appendix)
- $df_{group} = k - 1 = 3 - 1 = 2$ (numerator)
- $df_{error} = k(n - 1) = 3(4-1) = 9$ (denominator)
- $F_{.05}$ = ?

  4.26

F Distribution critical values (alpha = .05)

- $df_{group} = k - 1 = 3 - 1 = 2$ (numerator)
- $df_{error} = k(n - 1) = 3(4-1) = 9$ (denominator)

	Degrees of Freedom for Numerator									
df denom.	1	2	3	4	5	6	7	8	9	10
1	161.4	199.5	215.8	224.8	230.0	233.8	236.5	238.6	240.1	242.1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	$\boxed{4.26}$	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85

4. State Hypotheses

H_0: $\mu_1 = \mu_2 = \mu_3$ *referred to as the omnibus null hypothesis*

H_1: not all μ's are equal

- When rejecting the Null in ANOVA, we can only conclude that there is at least one significant difference among conditions.
- If ANOVA significant
  - pinpoint the actual difference(s), with <u>post-hoc comparisons</u>

Examine Data and Calculate F_{obt}

	X_1	X_2	X_3
	Control	Normal Cell	Hands-free Cell
DV: Response time (seconds)	.50	.75	.65
	.55	.65	.50
	.45	.60	.65
	.40	.60	.70

Is there at least one significant difference between conditions?

A Reminder about Variance

- Variance: the average squared deviation from the mean

$X = 2, 4, 6 \quad \overline{X} = 4$

Sample Variance Definitional Formula

$$s^2_X = \frac{\sum(X - \overline{X})^2}{N-1}$$

X	$X-\overline{X}$	$(X-\overline{X})^2$
2	-2	4
4	0	0
6	2	4
	$\sum(X - \overline{X})^2 =$	8

Sum of the squared deviation scores

$s^2_X = 8/2 = 4$

ANOVA Summary Table

Source	Sum of Squares	df	Mean Squares	F
Group	SS_{group}	df_{group}	MS_{group}	F_{obt}
Error	SS_{error}	df_{error}	MS_{error}	
Total	SS_{total}	df_{total}		

- SS = Sum of squared deviations or "Sum of Squares"

ANOVA Summary Table

Source	Sum of Squares	df	Mean Squares	F
Group	.072	df_{group}	MS_{group}	F_{obt}
Error	.050	df_{error}	MS_{error}	
Total	.122	df_{total}		

$SS_{total} = SS_{error} + SS_{group}$

$SS_{total} = .072 + .050 = .122$

ANOVA Summary Table

Source	Sum of Squares	df	Mean Squares	F
Group	.072	2	MS_{group}	F_{obt}
Error	.050	9	MS_{error}	
Total	.122	11		

- df between groups = k - 1
- df error (within groups) = $k\ (n - 1)$
- df total = N – 1
  (the sum of df_{group} and df_{error})

Examine Data and Calculate F_{obt}

- Compute the mean squares

$$MS_{group} = \frac{SS_{group}}{df_{group}} = \frac{.072}{2} = .0360$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{.050}{9} = .0056$$

Examine Data and Calculate F_{obt}

- Compute F_{obt}

$$F_{obt} = \boxed{\frac{MS_{group}}{MS_{error}}} = \frac{.0360}{.0056} = 6.45$$

ANOVA Summary Table

Source	Sum of Squares	*df*	Mean Squares	*F*
Group	.072	2	.0360	6.45
Error	.050	9	.0056	
Total	.122	11		

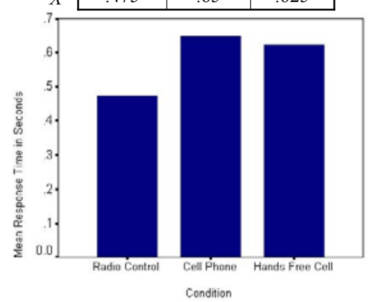
ANOVA Example: Cell phones

- Interpret results in terms of hypothesis
  6.45 > 4.26; Reject H_0 and accept H_1
- Report results
  $F(2, 9) = 6.45, \ p < .05$
- Explain in plain language
  – Among those three groups, there is at least one significant difference

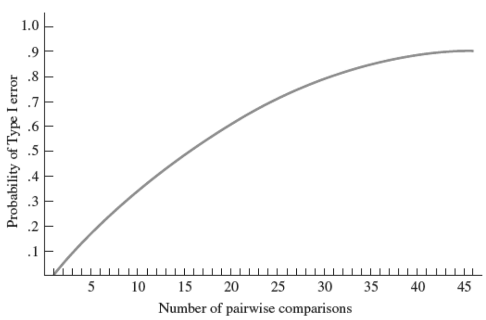
  – *more explanation is needed after post-hoc tests*

Interpret F_{obt}

$\overline{X}$	.475	.65	.625



- Figure 16.3 Probability of a Type I error as a function of the number of pairwise comparisons where $\alpha = .05$ for any one comparison



Post-hoc Comparisons

- Fisher's Least Signficant Difference Test (LSD)
  – uses *t*-tests to perform all pairwise comparisons between group means.
  – good with three groups, risky with > 3
  – this is a liberal test; i.e., it gives you high **power** to detect a difference if there is one, but at an increased risk of a Type I error

Post-hoc Comparisons

- Bonferroni procedure
  - uses *t*-tests to perform pairwise comparisons between group means,
  - but controls overall error rate by setting the error rate for each test to the familywise error rate divided by the total number of tests.
  - Hence, the observed significance level is **adjusted** for the fact that multiple comparisons are being made.
  - e.g., if six comparisons are being made (all possibilities for four groups), then alpha = .05/6 = .0083

Post-hoc Comparisons

- Tukey HSD
  (**H**onestly **S**ignificant **D**ifference)
  - sets the familywise error rate at the error rate for the collection for all pairwise comparisons.
  - very common test
- Other post-hoc tests also seen:
  - e.g., Newman-Keuls, Duncan, Scheffe'…

Effect Size: Partial Eta Squared

- Partial Eta squared (η^2) indicates the proportion of variance attributable to a factor
  - 0.20 small effect
  - 0.50 medium effect
  - 0.80 large effect
- Calculation: PASW

Effect Size: Omega Squared

- A less biased indicator of variance explained in the population by a predictor variable

$$\boxed{\omega^2 = \frac{SS_{\text{group}} - (k-1)MS_{\text{error}}}{SS_{\text{total}} + MS_{\text{error}}}}$$

$$\omega^2 = \frac{.072 - (3-1)(.0056)}{.122 + .0056} = 0.48$$

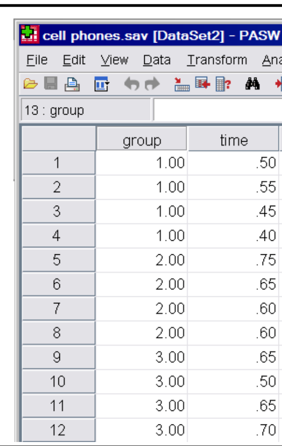
- 48% of the variability in response times can be attributed to group membership (medium effect)

PASW: One-Way ANOVA (Between Subjects)

- Setup a one-way between subjects ANOVA as you would an independent samples t-test:
- Create two variables
  - one variable contains levels of your independent variable (here called "group").
    - there are three groups in this case numbered 1-3.
  - second variable contains the scores of your dependent variable (here called "time")

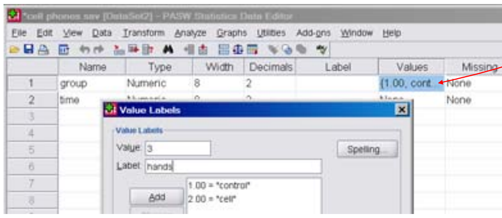
PASW :

One-Way ANOVA (Between Subjects)



	group	time
1	1.00	.50
2	1.00	.55
3	1.00	.45
4	1.00	.40
5	2.00	.75
6	2.00	.65
7	2.00	.60
8	2.00	.60
9	3.00	.65
10	3.00	.50
11	3.00	.65
12	3.00	.70

- **Label** the numbers you used to differentiate groups:
- Go to “Variable View”, then click on the “Values” box, then the gray box labeled “…”
- Enter Value (in this case 1, 2 or 3) and the Value Label (in this case: control, cell, hands)
- Click “Add”, and then add the next two variables.



Performing Test

- Select from **Menu: Analyze -> General Linear Model -> Univariate**
- Select your dependent variable (here: “time”) in the “Dependent Variable” box
- Select your independent variable (here: “group”) in the “Fixed Factor(s)” box
- Click “**Options**” button,
  - check **Descriptives** (this will print the means for each of your levels)
  - check **Estimates of effect size** for Partial Eta Squared
- Click the **Post Hoc** button for post hoc comparisons; move factor to “Post Hoc Tests for” box; then check “LSD, Bonferroni, or Tukey”
- Click **OK**

Descriptive Statistics

Dependent Variable:time

group	Mean	Std. Deviation	N
control	.4750	.06455	4
cell	.6500	.07071	4
hands	.6250	.08660	4
Total	.5833	.10517	12

if $p < .05$, then significant effect

Tests of Between-Subjects Effects

Dependent Variable:time

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.072[a]	2	.036	6.450	.018	.589
Intercept	4.083	1	4.083	735.000	.000	.988
group	.072	2	.036	6.450	.018	.589
Error	.050	9	.006			
Total	4.205	12				
Corrected Total	.122	11				

a. R Squared = .589 (Adjusted R Squared = .498)

Dependent Varia

Tukey HSD

control and cell groups are significantly different

(I) condition	(J) condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
control	cell	-.17500*	.05270	.022	-.3222	-.0278
	hands	-.15000*	.05270	.046	-.2972	-.0028
cell	control	.17500*	.05270	.022	.0278	.3222
	hands	.02500	.05270	.885	-.1222	.1722
hands	control	.15000*	.05270	.046	.0028	.2972
	cell	-.02500	.05270	.885	-.1722	.1222

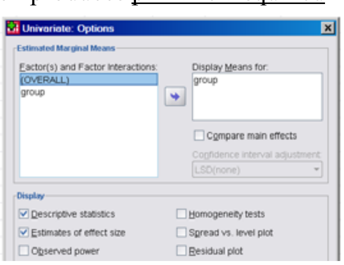
*. The mean difference is significant at the .05 level.

hands and cell groups are NOT significantly different

- Complete explanation
  - Any kind of cell phone conversation can cause a longer reaction time compared to listening to the radio.
  - There is no significant difference between reaction times in the normal cell phone and hands-free conditions.

PASW and Effect Size

- Click **Options** menu; then check **Estimates of effect size** box
- This option produces partial eta squared



Partial Eta Squared

Tests of Between-Subjects Effects

Dependent Variable:time

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.072[a]	2	.036	6.450	.018	.589
Intercept	4.083	1	4.083	735.000	.000	.988
group	.072	2	.036	6.450	.018	.589
Error	.050	9	.006			
Total	4.205	12				
Corrected Total	.122	11				

a. R Squared = .589 (Adjusted R Squared = .498)

Cell Phones

- Results: reporting and explanation
  - A one way, between subjects ANOVA with response time as the dependent variable showed a significant effect, $F(2, 9) = 6.45$, $p < .05$; Eta squared = .59. Tukey HSD post-hoc comparisons showed that the control radio group was significantly faster (M = .48, SD = .065) than the cell phone group (M = .65, SD = .071) and the hands-free cell phone group (M = .63, SD = .087). The cell phone group and hands-free group were not signficantly different.
  - These results suggest that any kind of cell phone conversation can cause a longer reaction time compared to listening to the radio. However, hands-free cell phones do not differ from regular cell phones.

PASW Data Example

- Three groups with three in each group ($N = 9$)

	Fast	Medium	Slow
	20.0	2.0	2.0
	44.0	22.0	2.0
	30.0	2.0	2.0
$\bar{X} =$	31.3	8.7	2.0

ANOVA Summary

Tests of Between-Subjects Effects

Dependent Variable:errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1418.667[a]	2	709.333	7.636	.022	.718
Intercept	1764.000	1	1764.000	18.990	.005	.760
groupSpeed	1418.667	2	709.333	7.636	.022	.718
Error	557.333	6	92.889			
Total	3740.000	9				
Corrected Total	1976.000	8				

a. R Squared = .718 (Adjusted R Squared = .624)

if $p < .05$, then significant effect

Effect size

Post Hoc Comparisons

slow and medium groups are *not* significantly different

Multiple Co

errors
Tukey HSD

(I) group Speed	(J) group Speed	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
slow	medium	22.6667	7.86930	.063	-1.4785	46.8118
	fast	29.3333*	7.86930	.023	5.1882	53.4785
medium	slow	-22.6667	7.86930	.063	-46.8118	1.4785
	fast	6.6667	7.86930	.690	-17.4785	30.8118
fast	slow	-29.3333*	7.86930	.023	-53.4785	-5.1882
	medium	-6.6667	7.86930	.690	-30.8118	17.4785

Based on observed means.
The error term is Mean Square(Error) = 92.889.

*. The mean difference is significant at the .05 level.

slow and fast groups are significantly different

Shamay-Tsoory SG, Tomer R, Aharon-Peretz J. (2005) The neuroanatomical basis of understanding sarcasm and its relationship to social cognition. *Neuropsychology*. 19(3), 288-300.

- A Sarcastic Version Item
  - Joe came to work, and instead of beginning to work, he sat down to rest.
  - His boss noticed his behavior and said, "Joe, don't work too hard!"
- A Neutral Version Item
  - Joe came to work and immediately began to work. His boss noticed his behavior and said, "Joe, don't work too hard!"
- Following each story, participants were asked:
  - Did the manager believe Joe was working hard?

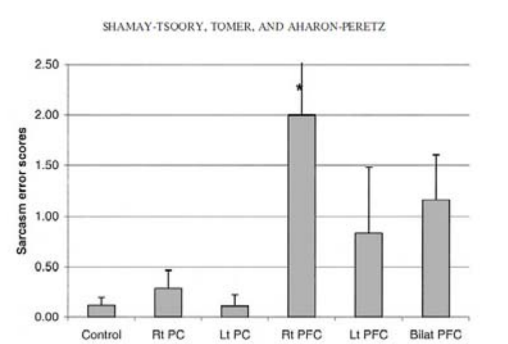


Figure 2. Mean sarcasm error scores (error bars represent standard error) in patients with unilateral lesions. Post hoc analysis: Right PFC was significantly different from left PFC, right PC, left PC, and controls (*$p < .05$). Rt = right; PC = posterior cortex; Lt = left; PFC = prefrontal cortex; Bilat = bilateral.